Londoño-Gil, M., Leite, N., Niehoff, T., Hanenberg, E., E. Grindflek, E. (2026) Validation of predicted Mendelian sampling variance for optimizing selection: an application in pigs, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23410: https://doi.org/10.31274/wcgalp.23410 ©2026 Londoño-Gil, et al. CC BY-NC-ND

Validation of predicted Mendelian sampling variance for optimizing selection: an application in pigs

M. Londoño-Gil[1*], N.G. Leite [1], T. Niehoff[2], E. Hanenberg[1], and E. Grindflek[1]

[1]Topigs Norsvin, Meerendonkweg 25, 5216 TZ 's-Hertogenbosch, The Netherlands;
*[2]Wageningen University & Research, Droevendaalsesteeg 1, P.O. Box 338, 6700 AH Wageningen, The Netherlands *Corresponding author: marisol.gil@topigsnorsvin.com*

Introduction

Modern animal breeding strives for genetic improvement while preserving genetic diversity. However, relying solely on breeding values in traditional selection can cause a decline in genetic variation over time. One way to address this is by considering the genetic variance of future offspring in selection decisions, emphasizing diversity through expected genetic gain. This can be achieved using predicted gametic Mendelian sampling variances (pred_MSVgam), which reflect an individual's contribution to the genetic variance of the next generation. The effectiveness of pred_MSVgam in predicting genetic diversity and improving long-term genetic response has been demonstrated (Santos *et al*., 2019). The gametic Mendelian Sampling Variance measures the variation in breeding values of gametes produced by an individual. It can be calculated for specific traits or for a selection index that combines multiple economically important traits. In this study, we aimed to (1) predict pred_MSVgam values for four pig finishing traits and a pseudo selection index and (2) validate these predictions with the Mendelian Sampling Variance observed in offspring (obs_MSVoff).

Materials & Methods

General dataset. The study used data from animals born between 2017 and 2024 of two terminal sire lines (P1 and P2) for traits including Average daily gain before testing (YDG), Average daily gain during testing (TDG), Backfat thickness at off-testing (BFE), and Loin Depth at off-testing (LDE). A total of 997,380 phenotypes, 1,050,924 pedigree records, and 410,877 genotypes (20,958 SNPs) were included in the complete data set. Based on the year of birth, a subset was created to simulate a scenario where the genotypes of the validation sires were available while their offspring's performance was not (partial data set). This setup reflects the information typically available at the selection point. Animals born after 2023 were removed from the full dataset, and the remaining data were used to form the partial subset, which contained 818,531 phenotypes, 334,986 genotypes, and 871,605 individuals with pedigree records. This subset was approximately 20% smaller than the complete set.

EBV and SNP effect estimation. A single-step GBLUP with a T factoring (ssGTBLUP, Vandenplas *et al*., 2023) multi-trait animal model, including the four traits of interest in this study, was used to obtain EBVs and SNP effects using the partial and complete data sets. The solutions were obtained with routine statistical models using the MiXBLUP v3.0.1 program (Vandenplas *et al*., 2022).

Gametic Mendelian Sampling Variance (pred_MSVgam). Gametic Mendelian Sampling Variance values were predicted using an internal tool based on the methodology described by Musa & Reinsch (2025). For this, phased genotypes, an internal genetic map containing SNP distances in Morgans (Brekke *et al*., 2022), and SNP effects, estimated from a partial data set, were used. Phased genotypes were obtained with Beagle 5.4 (Browning & Browning, 2007) using an in-house routine. Pred_MSVgam was predicted for each of the four traits for 34 sires

in P1 and 65 sires in P2. Afterward, pred_MSVgam was also calculated at the level of a pseudo index, which was used to simulate a real scenario where multiple traits are selected simultaneously.

Validation. To validate the pred_MSVgam values, genotyped sires with at least 100 phenotyped offspring for the four different traits were selected. These sires were parents of animals born after 2022. This threshold of 100 offspring per sire was chosen to ensure that the observed Mendelian Sampling Variance obtained from offspring information (obs_MSVoff) could be estimated with minimal influence from stochasticity due to the limited sample size. This represents a trade-off: increasing the minimum number of validation offspring per sire would reduce the impact of random errors in the variance estimate, but it would also decrease the number of sires meeting this requirement. Table 1 shows the number of validation sires used, categorized by line and trait.

Table 1. Number of validation sires with at least 100 phenotyped offspring in the validation set for the four traits of interest.

Trait	Sires P1	Sires P2
YDG	34 sires with 6,263 offspring	66 sires with 20,168 offspring
TDG	34 sires with 6,221 offspring	65 sires with 19,815 offspring
BFE	35 sires with 6,853 offspring	65 sires with 19,962 offspring
LDE	35 sires with 6,845 offspring	65 sires with 19,998 offspring

The variance of breeding values in the offspring generation (V_{BV}, Eq. 1) is influenced by the variance among breeding values of sires and dams (V_{sires}, V_{dams}), the covariance between them ($COV_{sires,dams}$), which is non-zero in assortative mating, as well as the gametic Mendelian Sampling Variances of sires and dams ($\bar{V}_{MS\ sires}$, $\bar{V}_{MS\ dams}$). The sum of the first three terms is the variance of parent averages in the next generation.

$$V_{BV} = \frac{V_{sires}}{4} + \frac{V_{dams}}{4} + \frac{COV_{sires,dams}}{2} + \bar{V}_{MS\ sires} + \bar{V}_{MS\ dams} \quad (1)$$

The variance of offspring was calculated for each paternal half-sib progeny group, which removes the terms V_{sires} and $COV_{sires,dams}$. Validation of the gametic MSV was performed by correlating pred_MSVgam of the validation sires with the obs_MSVoff based on their offspring born between 2023 and 2024, where obs_MSVoff is the variance of estimated Mendelian Sampling terms in the half-sib progeny group. To account for the possibility that the variance among breeding values of dams (V_{dams}) may differ between half-sib groups, obs_MSVoff was obtained as in Eq. 2:

$$\text{obs}_{\text{MSVoff}} = var(MS_i) = var\left(EBV_i - \left(\frac{EBVsire_i + EBVdam_i}{2}\right)\right) \quad (2)$$

Thus, obs_MSVoff is the variance of estimated MS terms in the progeny group. This property differs from pred_MSVgam, which is based solely on the MS terms provided by the sire. However, this validation scheme was adopted because separating the paternal and maternal MS terms is challenging. Based on pred_MSVgam values, the two most extreme sires were classified into bottom (i.e., sire with the lowest pred_MSVgam) and top (i.e., sire with the highest pred_MSVgam). Distributions of the MS_i for their progeny were then plotted to assess concordance with expectations. Furthermore, correlations between pred_MSVgam and obs_MSVoff for sires were calculated across sires.

Results

Fig. 1 shows the distribution of MS_i of offspring of validation sires assigned as top and bottom for each of the four traits evaluated in P1, while Fig. 2 shows this distribution at the index level. Results for individual traits are not shown for P2 (similar patterns).

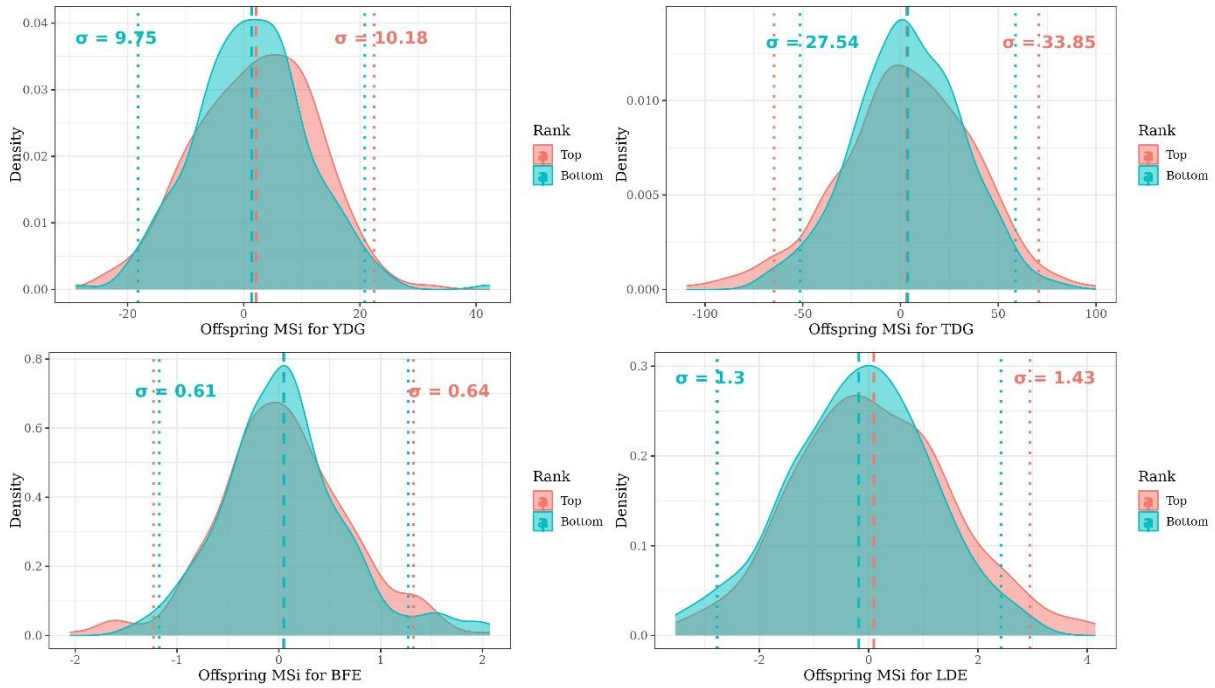

Figure 1 Distributions of MS_i of offspring of validation sires classified as top or bottom based on pred_MSVgam for YDG (top-left), TDG (top-right), BFE (bottom-left), and LDE (bottom-right) in P1. The dashed central lines mark the mean MS_i and the dotted lateral lines represent the 95% interval for each group. σ indicates the observed standard deviation of the MS_i.

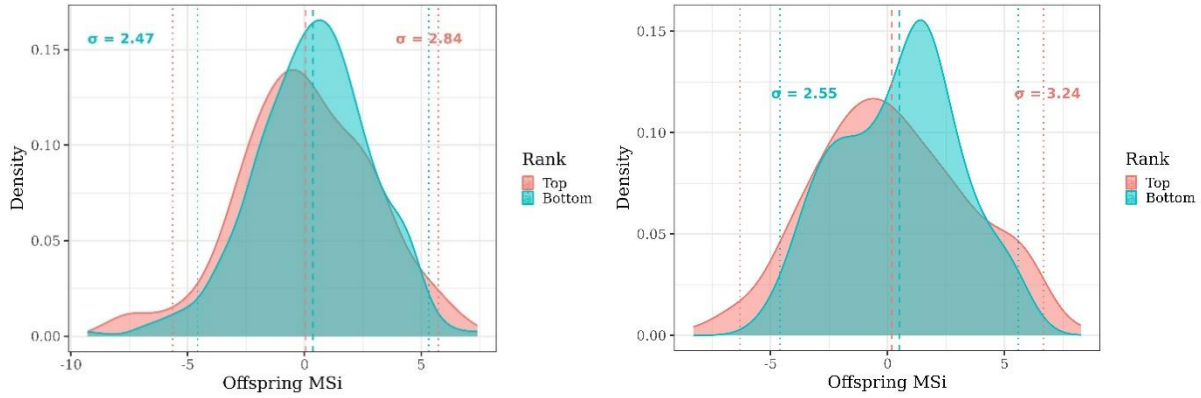

Figure 2. Distributions of MS_i of offspring of validation sires classified as top or bottom based on pred_MSVgam over the selection index for P1 (left) and P2 (right). The dashed central lines mark the mean MS_i and the dotted lateral lines represent the 95% interval for each group. σ indicates the observed standard deviation of the MS_i.

Correlation between pred_MSVgam and obs_MSVoff ranged from 0.20 to 0.63 for P1 and from 0.51 to 0.68 for P2 for individual traits (Table 2). Fig. 3 shows the correlations (scatter plots) between obs_MSVoff and pred_MSVgam for the selection index.

Table 2. Correlations between pred_MSVgam and obs_MSVoff in the P1 and P2 populations.

Trait	P1	P2
YDG	0.201	0.505
TDG	0.628	0.640
BFE	0.457	0.684
LDE	0.508	0.606

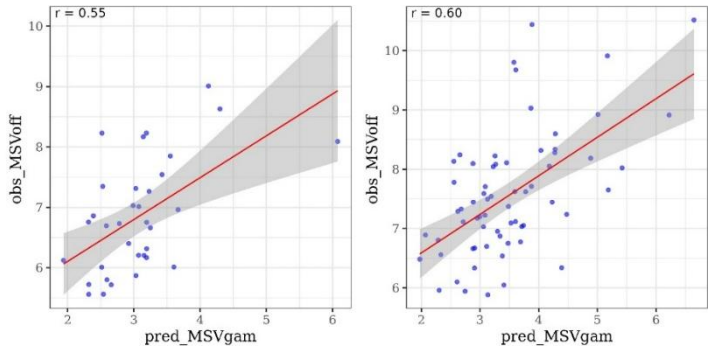

Figure 3. Scatter plots of obs_MSVoff versus pred_MSVgam for the selection index in P1 (left) and P2 (right). Blue dots: individual sires, red line: fitted linear regression. r: Pearson correlation.

Discussion

Results shown here indicate that sires ranked highest for pred_MSVgam produced progeny with greater genetic variability (Fig. 1). This finding was consistent at both the individual trait level (Fig. 1) and the index level (Fig. 2) across both populations. These figures demonstrate that individuals with high gametic MSV generate more variable progeny, thereby increasing the selection differential. The predictability of Mendelian Sampling Variance in offspring varied between traits and lines. Notably, P2 showed slightly stronger correlations than P1 at both the individual trait and index levels (Fig. 3). These differences may be due to P2's larger reference population size and differences in sample size per sire. For example, sires in P1 had an average of 196 offspring (6,853/35), while sires in P2 had an average of 306 offspring (20,168/66). Consequently, variance estimates in P2 are based on larger progeny groups and are less influenced by random sampling errors. Theoretical expectations support this interpretation. When assuming no estimation error in SNP effects or breeding values, the maximum achievable correlation with this validation scheme is roughly 0.70 for P1 and 0.77 for P2. This difference reflects the larger offspring group size in P2, which further diminishes the impact of stochasticity on observed variances. Overall, the observed correlations, especially at the index level, are promising and indicate that predicted offspring variances could be incorporated into future selection criteria, as suggested in other studies (Bijma *et al*., 2020; Niehoff *et al*., 2024; Santos *et al*., 2019).

Conclusion

pred_MSVgam is a useful predictor of genetic variability in progeny, exhibiting consistent, moderate-to-strong associations across traits and populations. These results support its potential use to improve selection by exploiting genetic variability and indicate that predicted offspring variances could be incorporated into future selection criteria.

References

Bijma P., Wientjes Y.C., and Calus M.P. (2020) Genetics 214:91-107. https://doi.org/10.1534/genetics.119.302643
Brekke C., Berg P., Gjuvsland A.B., *et al*. (2022) Genet. Sel. Evol. 54:33. doi.org/10.1186/s12711-022-00723-9
Browning, S. R., and Browning, B. L. (2007) Am. J. Hum. Genet. 81:1084-1097. doi.org/10.1086/521987
Musa A.A., and Reinsch N. (2025) J. Anim. Breed. Genet. doi.org/10.1111/jbg.12930
Niehoff T.A., Ten Napel J., Bijma P., *et al*. (2024) Genet. Sel. Evol. 56:41. doi.org/10.1186/s12711-024-00899-2
Santos D.J.A., Cole J.B., Lawlor Jr T.J., *et al*. (2019) J. Dairy Sci. 102:5279-5294. doi.org/10.3168/jds.2018-15971
Vandenplas, J., Ten Napel, J., Darbaghshahi, S. N., *et al*. (2023) Genet. Sel. Evol. 55:37. doi.org/10.1186/s12711-023-00808-z
Vandenplas J., Veerkamp R.F., Calus M.P.L., *et al*. (2022) Proc. 12th WCGALP, Rotterdam, Netherlands